MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2017/2018

DPS5018 – INTRODUCTION TO PROBABILITY AND STATISTICS

(T41-49, T51-53 , T55-57, 5055 , 6065)

26 OCTOBER 2017
9.00 A.M. – 11.00 A.M.
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This Question paper consists of 4 pages with 5 questions, excluding cover page and appendix.

2. Attempt ALL **FIVE** questions.

3. Write all your answers in the Answer Booklet provided.

4. Key formulae are given in the Appendix.

Question 1

a. "*A sales promoter decides to visit 10 houses in a street and offer special discounts on a new window treatment. The street has 100 houses numbered from 1 to 100. The sales promoter selects a random number between 1 and 10 inclusive and calls on the house with that street number. After that the promoter calls on every tenth house.*"
From the above information, identify the
   i. population and sample. (2 marks)
   ii. sampling technique used. (1 mark)

b. In each of the statements below, identify whether descriptive or inferential statistics have been used. (3 marks)
   i. The median household income for people aged 25 to 34 is RM2500.
   ii. Nine out of ten on the job fatalities are men.
   iii. Based on previous study, drinking decaffeinated coffee can raise cholesterol level by 7%.

c. Classify each variable as quantitative or qualitative. (4 marks)
   i. Classification of children in Bukit Beruang Day-Care Centre either infant, toddler or preschool.
   ii. Weight of fish caught in Sungai Putat Melaka.
   iii. Capacity of a cubic feet of six truck bed.
   iv. Marital status of faculty members in MMU.

[TOTAL 10 MARKS]

Continued...

Question 2

a. The following data shows the marks obtained by 25 Diploma students in Statistics Test.

42	67	57	67	45
81	50	65	51	54
55	61	43	76	63
84	55	62	52	77
49	72	58	73	59

i. Represent the results above by using stem-and-leaf diagram. (3 marks)

ii. List out the five number summary. (5 marks)

iii. Hence, find the range and interquartile range. (2 marks)

b. The table below shows a sample of height of 70 students in MMU.

Height (cm)	Number of students	Midpoint, x	x^2	fx	fx^2
$145 < 150$	7				
$150 \leq 155$	12				
$155 < 160$	20				
$160 \leq 165$	14				
$165 \leq 170$	9				
$170 < 175$	6				
$175 \leq 180$	2				
				$\sum fx =$	$\sum fx^2 =$

From the above information,

i. reconstruct and complete the table in the answer booklet provided. (5 marks)

ii. calculate the sample mean. (2 marks)

iii. find the sample variance. (3 marks)

[TOTAL 20 MARKS]

Continued...

Question 3

a. The probability of vegetables selected for sale from farm A is $\frac{3}{4}$ while the probability of vegetables selected for sale from farm B is $\frac{7}{9}$. Find the probability that the vegetables selected for sale are only from one farm. (3 marks)

b. The insurance company has 100 salespersons, 40% of them are male and the rest are female. Sixteen of the male person are bachelors. Of all the female salespersons, 36 of them are married. A salesperson is selected at random from the company.

i. Display the above information in a contingency table. (3 marks)

ii. It is known that the salesperson selected is single, what is the probability that the salesperson is a male? (3 marks)

c. A traffic engineer is interested in the number of vehicles reaching a particular crossroad during periods of relatively low traffic flow. The engineer finds that the number of vehicles X reaching the cross roads per minute is governed by the probability distribution:

x	0	1	2	3	4
$P(X=x)$	0.37	0.39	0.19	0.04	0.01

Calculate the

i. expected value (2 marks)

ii. standard deviation (3 marks)

of the random variable X.

d. It is found that 60% of the students from a certain class obtained grade A in Statistics subject. If 7 students from the class are selected at random, find the probability that

i. exactly 3 students obtained grade A. (2 marks)

ii. not more than 5 students obtained grade A. (2 marks)

e. An insurance salesperson sells an average of 1.4 policies per day. Find the probability that on a certain day, this salesperson will sell

i. less than 3 insurance policies. (2 marks)

ii. more than 10 insurance policies in one week. (3 marks)

f. The masses of oranges selling at a fruit stall follow a normal distribution with a mean of 3.5 kg and a standard deviation of 0.4 kg. find

i. the probability that an orange chosen randomly from the fruit stall has a mass of more than 4.2 kg. (3 marks)

ii. the value of b if 70% of the oranges from the fruit stall with mass of more than b kg. (4 marks)

[TOTAL 30 MARKS]

Continued...

Question 4

The data below are working hours of workers from eight different construction companies and their salaries.

Working hours, x	Salaries (RM), y
14	2500
12	2200
10	2000
9	1900
11	2150
13	2350
7	1750
8	1500

a. Calculate the value of $\sum xy$. (3 marks)

b. Show that $\sum x^2 = 924$. (2 marks)

c. Given that the values $\sum x = 84$, $\sum y = 16350$ and $\sum y^2 = 34157500$, compute the values of SS_{XX}, SS_{XY} and SS_{YY}. (6 marks)

d. Find the regression equation. (6 marks)

e. Determine the value of linear correlation, r and interpret your answer. (3 marks)

[TOTAL 20 MARKS]

Question 5

a. A previous study by the MMU Student Council claims that students spend on average of 9.5 hours a day on social media. A new survey has been conducted recently to test this claim. A random sample of 62 students was selected and it is found that they spend on average of 7.8 hours and a standard deviation of 5.35 hours a day on social media. Is there any evidence to reject the claim by the MMU Student Council? Test at 10% level of significance by using p–value approach. (9 marks)

b. Customer service department of Good Recipe Company wanted to compare the preference of cake flavor between male and female customer. They want to check whether the demand for chocolate cake flavor among male customer is higher compared to female customer. To check this claim, they choose a simple random sample of 120 male and 105 female customers. They found that 66 of male customer and 42 of female customer demand chocolate cake flavor. By using p–value approach, what can they conclude based on these findings? Test at 3% level of significance. (11 marks)

[TOTAL 20 MARKS]

End of Page.

APPENDIX – KEY FORMULAE

Measures of Central of Tendency and Measures of Dispersion

SAMPLE	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{\sum x_i f_i}{n}$
Variance	$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$	$s^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n-1}$

** For grouped data, x is the midpoint and f is the frequency of the class.

POPULATION	Ungrouped Data	Grouped Data
Mean	$\mu = \frac{\sum x_i}{N}$	$\mu = \frac{\sum x_i f_i}{N}$
Variance	$\sigma^2 = \frac{\sum x_i^2}{N} - \mu^2$	$\sigma^2 = \frac{\sum f x_i^2}{N} - \mu^2$

** For grouped data, x is the midpoint and f is the frequency of the class.

Discrete Random Variable

Mean:	Standard deviation:
$\mu = \sum xP(X = x)$	$\sigma = \sqrt{\sum x^2 P(X = x) - \mu^2}$

Binomial Distribution

Binomial probability formula:	$P(X = x) = \binom{n}{x} p^x q^{n-x}$

Key formula for Binomial Probability Distribution

If $p \leq 0.5$,

i. $P(X = r) = B(r) - B(r-1)$
ii. $P(X \geq r) = 1 - B(r-1)$
iii. $P(X \leq r) = B(r)$
iv. $P(X > r) = 1 - B(r)$
v. $P(X < r) = B(r-1)$
vi. $P(a \leq X \leq b) = B(b) - B(a-1)$

If $p > 0.5$	i.	$P(X = r) = B(n - r) - B(n - r - 1)$
	ii.	$P(X \geq r) = B(n - r)$
	iii.	$P(X \leq r) = 1 - B(n - r - 1)$
	iv.	$P(a \leq X \leq b) = B(n - a) - B(n - b - 1)$

Poisson Distribution

Poisson probability formula:	$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

Key formula for Poisson Probability Distribution:

If $\mu = \lambda$,

	i.	$P(X = r) = Poi(r) - Poi(r - 1)$
	ii.	$P(X \geq r) = 1 - Poi(r - 1)$
	iii.	$P(X \leq r) = Poi(r)$
	iv.	$P(X > r) = 1 - Poi(r)$
	v.	$P(X < r) = Poi(r - 1)$
	vi.	$P(a \leq X \leq b) = Poi(b) - Poi(a - 1)$

Normal Probability Distribution

z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

Linear Regression

For the **least square regression**, $\hat{y} = a + bx$

where $SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$, $SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$

and $b = \frac{SS_{xy}}{SS_{xx}}$, $a = \bar{y} - b\bar{x}$

Hypothesis Testing

	Mean	Proportion
One Population	$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{p - p_0}{\sqrt{p_0(1-p_0)/n}}$
Two Population	$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$z = \frac{(p_1 - p_2)}{\sqrt{p_p(1-p_p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ where $p_p = \frac{x_1 + x_2}{n_1 + n_2}$